

# UC Berkeley

## UC Berkeley Previously Published Works

### Title

Using microarrays to facilitate positional cloning: identification of tomosyn as an inhibitor of neurosecretion.

### Permalink

<https://escholarship.org/uc/item/60n8p74d>

### Journal

PLoS genetics, 1(1)

### ISSN

1553-7390

### Authors

Dybbs, Michael  
Ngai, John  
Kaplan, Joshua M

### Publication Date

2005-07-01

### DOI

10.1371/journal.pgen.0010002

Peer reviewed

# Using Microarrays to Facilitate Positional Cloning: Identification of Tomosyn as an Inhibitor of Neurosecretion

Michael Dybbs<sup>1,2</sup>, John Ngai<sup>2</sup>, Joshua M. Kaplan<sup>1\*</sup>

**1** Department of Molecular Biology, Massachusetts General Hospital, Boston, Massachusetts, United States of America, **2** Department of Molecular and Cell Biology, Functional Genomics Laboratory, Helen Wills Neuroscience Institute, University of California, Berkeley, California, United States of America

**Forward genetic screens have been used as a powerful strategy to dissect complex biological pathways in many model systems. A significant limitation of this approach has been the time-consuming and costly process of positional cloning and molecular characterization of the mutations isolated in these screens. Here, the authors describe a strategy using microarray hybridizations to facilitate positional cloning. This method relies on the fact that premature stop codons (i.e., nonsense mutations) constitute a frequent class of mutations isolated in screens and that nonsense mutant messenger RNAs are efficiently degraded by the conserved nonsense-mediated decay pathway. They validate this strategy by identifying two previously uncharacterized mutations: (1) *tom-1*, a mutation found in a forward genetic screen for enhanced acetylcholine secretion in *Caenorhabditis elegans*, and (2) an apparently spontaneous mutation in the *hif-1* transcription factor gene. They further demonstrate the broad applicability of this strategy using other known mutants in *C. elegans*, *Arabidopsis*, and mouse. Characterization of *tom-1* mutants suggests that TOM-1, the *C. elegans* ortholog of mammalian tomosyn, functions as an endogenous inhibitor of neurotransmitter secretion. These results also suggest that microarray hybridizations have the potential to significantly reduce the time and effort required for positional cloning.**

Citation: Dybbs M, Ngai J, Kaplan JM (2005) Using microarrays to facilitate positional cloning: Identification of tomosyn as an inhibitor of neurosecretion. PLoS Genet 1(1): e2.

## Introduction

Forward genetic screens have been traditionally utilized in model systems (e.g., *Caenorhabditis elegans*, *Drosophila*, yeast, and *Arabidopsis*). More recently, large-scale screens have been undertaken in vertebrate systems such as zebrafish [1,2] and mouse [3–5]. Mutations isolated in genetic screens are typically identified by positional cloning. The difficulty posed by positional cloning is determined by the size of the genome, the recombination rate, and the difficulty of assessing the mutant phenotype. For example, the mouse genome comprises 3,600 centimorgans (cM) and  $3 \times 10^9$  base pairs. The ultimate goal of a typical positional cloning project is to analyze a sufficient number of recombinants to map the mutation to a small genetic interval (typically approximately 0.1 cM). Once a mutation has been precisely mapped, gene identification is typically achieved by a variety of strategies: direct sequencing of the region (100 kb in the mouse), candidate gene testing, or screening for informative alleles (e.g., microdeletions). The difficulty of a particular positional cloning can be compounded by the nature of the mutant phenotype. This problem is particularly acute for behavioral mutants, which often have phenotypes that must be scored in multiple trials, or in populations of animals. Together, these issues conspire to make traditional positional cloning a significant and costly bottleneck.

To circumvent these difficulties, several new technologies have been developed to isolate mutations by reverse genetics. Reverse genetic strategies include use of insertional mutagens [6–10], PCR screens for randomly induced deletions [11], homologous gene targeting [12,13], and physical or genetic detection of point mutations in sequenced genes [14,15].

While reverse genetic strategies circumvent the positional cloning bottleneck, these approaches also have limitations. Mutations isolated by reverse genetics often lack obvious phenotypic defects (e.g., because they are in functionally redundant genes). Phenotypic differences observed in mutants isolated by reverse genetics can be confounded by other mutations in the genetic background, particularly since animals are typically heavily mutagenized in these strategies. For these reasons, it would be useful to develop methods that would allow more rapid characterization of mutations isolated in forward genetic screens.

We wondered whether microarray expression data could facilitate the identification of mutations responsible for behavioral defects isolated in forward genetic screens. It is well established that nonsense mutations result in the degradation of the mutant messenger RNA (mRNA) via the nonsense-mediated decay (NMD) pathway. A surveillance mechanism common to all eukaryotes, NMD serves as a quality control system to destroy faulty mRNAs whose

Received January 13, 2005; Accepted February 1, 2005; Published July 25, 2005  
DOI: 10.1371/journal.pgen.0010002

Copyright: © 2005 Dybbs et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abbreviations: ACh, acetylcholine; CaMKII, type II calcium- and calmodulin-dependent protein kinase; cM, centimorgan; mRNA, messenger RNA; NMD, nonsense-mediated decay

Editor: Gregory S. Barsh, Stanford University School of Medicine, United States of America

\*To whom correspondence should be addressed. E-mail: kaplan@molbio.mgh.harvard.edu

## Synopsis

Genetic screens are commonly used to figure out which genes are involved in a biological process. The first step in a genetic screen is to isolate mutant animals that are defective in the process being studied. The next step is to find which of the thousands of genes has the mutation that causes the observed defect. Positional cloning, the tried-and-true method for locating mutations, is slow and expensive. The authors propose using microarray hybridizations to speed the process. Their approach relies on the fact that a large fraction of the mutations found in screens are the results of premature stop codons, a particularly severe type of mutation. In cells, messages containing premature stop codons are rapidly destroyed by a protective pathway, called nonsense-mediated decay, thus making them directly detectable by microarray hybridization.

The authors apply this strategy retrospectively to known mutants in *Caenorhabditis elegans*, *Arabidopsis*, and mouse. They identify two uncharacterized mutations in *C. elegans*, including one, *tom-1*, found in a forward genetic screen for enhancers of neurotransmission. Interestingly, their characterization of *tom-1* mutants suggests that the highly conserved protein tomosyn inhibits neurotransmission in neurons. This study shows that microarray hybridizations will help reduce the time and effort required for positional cloning.

translation would lead to an inappropriately truncated protein [16–18]. NMD protects cells by eliminating inactive or potentially deleterious dominant negative proteins that are the result of somatic mutation, transcriptional mistakes, or splicing errors.

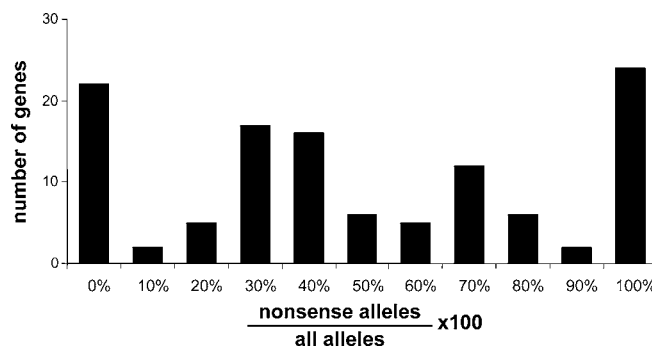
It has been proposed that NMD could be used as a basis to identify nonsense mutations in cell lines [19,20]. In principle, a nonsense mutation in mutant animals could be identified using microarray hybridizations to find transcripts with decreased abundance. In practice, microarray data alone are unlikely to be sufficient to identify nonsense mutations. In addition to the expected statistical noise associated with microarray experiments, there are likely to be transcriptional changes in other genes that are caused by the mutation being studied. The most powerful cloning approach would thus be one that uses microarray data together with traditional mapping information. Here, we present evidence supporting the feasibility and general utility of this strategy.

## Results

To test the feasibility of using microarrays to facilitate positional cloning, we will address four questions. (1) How frequently are nonsense alleles recovered in forward genetic screens? (2) Are microarray hybridizations sensitive enough to detect the decreased abundance of a nonsense mutant transcript? (3) Can microarray hybridizations be used to identify an uncloned behavioral mutant in *C. elegans*? (4) Is this microarray-based strategy applicable to other model organisms?

### Nonsense Alleles Represent a Large Fraction of *C. elegans* Mutations

The utility of microarrays in cloning depends on the frequency with which nonsense alleles are recovered in phenotypic screens. Since 15 of the 61 amino acid-encoding codons are mutable to stop codons by a single base-pair



**Figure 1.** Fraction of *C. elegans* Alleles That Are Nonsense Mutations

Molecular information about alleles was obtained from WormBase and literature searches. Graph includes 117 genes for which molecular characterization of three or more alleles was available (770 alleles total). Of these 117 alleles, 22 (19%) have no known nonsense mutations. Many of these no-nonsense alleles are in genes that are required for viability. DOI: 10.1371/journal.pgen.0010002.g001

substitution, nonsense alleles are likely to represent a large fraction of all alleles recovered after random mutagenesis with agents that increase the rate of nucleotide misincorporation. To assess the prevalence of nonsense alleles isolated following random mutagenesis, we compiled a list of sequenced *C. elegans* mutant alleles by downloading information from WormBase and conducting targeted literature searches (Figure 1; Table S1). We focused on misincorporation mutations because these represent the most common lesion caused by alkylating agents such as ethyl methane sulfonate or N-ethyl-N-nitrosourea. In total, we examined 943 single-nucleotide-substitution loss-of-function alleles in 246 genes, which were published by 99 laboratories. We then classified these alleles as putative NMD targets (nonsense) or non-NMD targets (missense). We found that 41% of the alleles were putative NMD targets. Interestingly, this figure is comparable to estimates that one-third of human genetic diseases are the result of nonsense mutations [21,22].

We calculated the percentage of nonsense alleles recovered for each of the 117 genes in our dataset with three or more characterized alleles (Figure 1). For 20% of these genes, no nonsense alleles have been identified, even in cases where ten alleles are known. Many of these genes are known to cause a lethal phenotype as a null mutation (e.g., *acy-1*, *unc-17*, and *let-502*). Since many screens demand homozygous viable phenotypes, it is not surprising that nonsense alleles were rarely recovered in these genes. For another 20% of these genes, all known alleles are nonsense mutations. These might comprise genes for which mutant phenotypes are expressed only when gene function is completely eliminated. For all other genes, there appears to be a broad distribution in the fraction of nonsense alleles recovered, with a mode occurring at 40% nonsense alleles. Thus, while a high frequency of nonsense alleles seems to be a general feature common to many screens in *C. elegans*, this is likely to vary considerably for different gene classes.

### Proof of Principle: *mec-3* and *unc-43* CaMKII Mutations Are Detectable by Microarray

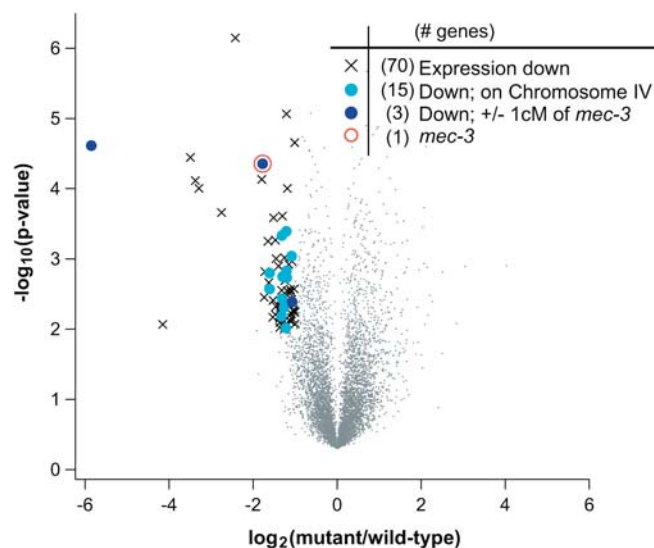
Are microarray hybridizations sensitive enough to detect changes in mutant transcript abundance due to a nonsense lesion above the global variation in gene expression between

mutant and control strains? Some potential sources of variance in gene expression include random fluctuations in gene expression [23,24], uncontrolled differences between the mutant and control populations (e.g., differences in developmental stage or physiological status), and differences in genetic backgrounds [25,26]. Perhaps the most important potential limitation is changes in gene expression that are a secondary consequence of a mutation. This could be particularly problematic for mutations in genes encoding transcription factors or other components of signal transduction cascades, the loss of which would be expected to alter the expression of many downstream genes.

To address some of these concerns, we examined the large collection of microarray experiments used to build a whole-genome expression profile for *C. elegans* [27]. Most of these experiments, which were done with printed microarrays, were designed to identify gene-expression profiles associated with various developmental programs or specific tissues. However, one set of experiments analyzed changes in gene expression in mutants lacking the MEC-3 transcription factor (see Materials and Methods) [28]. The *mec-3(e1338)* allele corresponds to a W69Stop mutation, and homozygous animals carrying this mutation are touch-insensitive [29,30].

Using this dataset, we classified genes as differentially expressed in *mec-3(e1338)* based on two criteria: average fold-change in expression level and statistical significance using a Student's *t*-test. We constructed a volcano plot with the  $\log_2(\text{fold-change})$  on the *x*-axis and negative  $\log_{10}(p\text{-value})$  on the *y*-axis [31]. This provides a useful way to visualize differentially expressed genes—those whose expression level is down (negative on the *x*-axis) and that show high statistical significance (large on the *y*-axis). Seventy genes were identified as having significantly reduced expression in *mec-3(e1338)*, using fold-change greater than  $-1.0$  ( $\log_2$  scale) and  $p < 0.01$  as thresholds for decreased expression (Figure 2). Had *e1338* been an uncharacterized mutation that we were attempting to clone, the next step would be to narrow the candidate list of 70 genes using mapping data. Fifteen of these genes are on Chromosome 4, which contains *mec-3* and approximately 2,900 other genes. Of these, only three differentially expressed genes fall within a two-map-unit interval spanning *mec-3* and approximately 100 other genes. Thus, even in the case of a transcription factor, microarray hybridizations are sufficiently sensitive to detect changes in the mutant mRNA abundance despite broader changes in gene expression. In the case of MEC-3, it is likely that the reduced abundance of *e1338* mRNA is due both to NMD and to positive autoregulation of *mec-3* transcription [32]. Therefore, triangulating between rough mapping information and microarray data would have facilitated the rapid cloning of *mec-3*.

To further address the sensitivity of microarray-assisted cloning, we analyzed changes in gene expression observed in KP3365 *unc-43(n1186)* mutants. The *unc-43* gene encodes type II calcium- and calmodulin-dependent protein kinase (CaMKII), which is broadly expressed in the worm nervous system as well as in muscles and in the intestine [33]. This provides another demanding test case because CaMKII plays a pivotal role in calcium-mediated signaling in neurons, and *unc-43* mutations are known to cause changes in the expression of other genes [34]. The *n1186* allele corresponds to a Q67Stop



**Figure 2.** Analysis of mRNA Abundance in *mec-3(e1338)* Animals

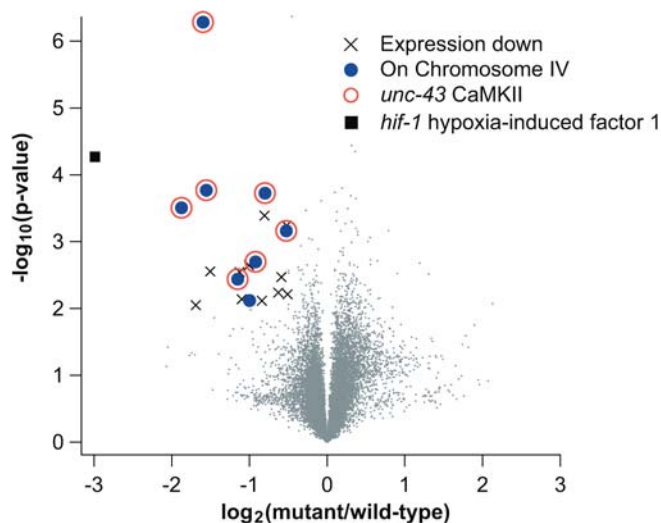
Fold-change (*x*-axis) is plotted against the statistical significance (*y*-axis) for each probeset. Fold-changes are shown on  $\log_2$  scale. *p*-Values are shown on a negative  $\log_{10}$  scale. The symbol  $\times$  indicates genes with reduced expression in *mec-3(e1338)* animals (fold-change  $< -1$ ,  $p < 0.01$ ). Light blue circles indicate genes with reduced expression that are also on Chromosome 4. Dark blue circles indicate genes with reduced expression that are within 1 cM to the left or right of *mec-3*. The open red circle indicates the *mec-3* gene.

DOI: 10.1371/journal.pgen.0010002.g002

mutation, and homozygous animals carrying this mutation have relatively subtle behavioral defects [33].

We hybridized total RNA isolated from wild-type and KP3365 *unc-43(n1186)* CaMKII mutant animals to the Affymetrix *C. elegans* GeneChip (Dataset S1). Using fold-changes greater than 0.5 ( $\log_2$  scale) and  $p < 0.01$  as thresholds, we found 20 probesets with decreased expression in KP3365 *unc-43(n1186)* CaMKII mutants as compared to wild-type controls (Figure 3). Eight of these probesets correspond to sequences on Chromosome 4, which contains *unc-43* CaMKII and approximately 2,900 other genes. Strikingly, seven of these eight probesets correspond to the *unc-43* gene. These seven probesets correspond to nonoverlapping regions of the coding sequence, as well as the 5' and 3' untranslated region of *unc-43* (Figure S1). Only two of these probesets (193459\_s\_at and 193463\_s\_at) were annotated as corresponding to the *unc-43* CaMKII mRNA transcript according to the annotation of the *C. elegans* GeneChip provided by Affymetrix (downloadable at <http://www.affymetrix.com>). During our examination of the eight candidate probesets that showed decreased expression and were on Chromosome 4, we discovered the five additional probesets that corresponded to *unc-43* CaMKII. These additional probesets provided a serendipitous blind control, since we were not aware of their existence until they appeared on our candidate list from the hybridization. This redundancy in probes is a result of overlap in the various databases used to design the GeneChip and inaccuracies in gene predications at the time of design (December 2000). Approximately 2,400 (13%) of the genes on the *C. elegans* GeneChip are represented by multiple probes. While it is unusual to have seven probesets on a microarray corresponding to a single gene, this illustrates the robustness of our ability to measure the changes in





**Figure 3.** Analysis of mRNA Abundance in the KP3365 *unc-43(n1186)* CaMKII Strain

Expression data are illustrated as described in Figure 2. The symbol  $\times$  indicates probesets with reduced expression in KP3365 animals (fold-change  $< -0.5$ ,  $p < 0.01$ ). Filled blue circles indicate probesets with reduced expression that are also on Chromosome 4. Open red circles indicate probesets corresponding to the *unc-43* CaMKII gene. The black square indicates the probe corresponding to the *hif-1* gene. DOI: 10.1371/journal.pgen.0010002.g003

expression of *unc-43* CaMKII. This demonstrates that microarray hybridizations can be used to measure the decrease in *unc-43* transcript in the *unc-43(n1186)* mutant. These results also suggest that microarray hybridizations utilizing printed arrays and those utilizing Affymetrix chips are both sufficiently sensitive to detect changes in the abundance of mutant mRNAs, even in the relatively stringent cases of transcription factors and signal transduction components.

#### Identification of a *hif-1* Polymorphism in the KP3365 Strain

One potential limitation of our strategy is that mutant strains may contain multiple mutations, some of which do not contribute to the mutant phenotype. This will be particularly true in heavily mutagenized strains, and in cases where the mutants have not been extensively backcrossed with wild-type strains. Therefore, we examined the KP3365 *unc-43(n1186)* CaMKII hybridization data for other genes with significantly reduced expression. Interestingly, the gene with the largest decrease in expression in KP3365 *unc-43(n1186)* animals was not *unc-43*; rather, it was *hif-1* (Figure 3), which encodes the worm ortholog of hypoxia-inducible factor 1 $\alpha$ , a transcription factor that mediates transcriptional responses to oxygen deprivation [35]. This is the only probe corresponding to *hif-1* on the *C. elegans* GeneChip. There are two likely explanations for the observed decrease in *hif-1* transcript levels: either *hif-1* expression is regulated by *unc-43* CaMKII, or the KP3365 strain contains a loss-of-function polymorphism in the *hif-1* gene. Sequencing genomic DNA from KP3365 animals revealed two mutations in the last exon of *hif-1* (*nu469*) (Figure 4A). Since neither of these mutations results in a premature stop, why is transcript level decreased? To address this issue, we sequenced *hif-1* cDNA made from wild-type and KP3365 animals. This revealed that the *nu469* mutations cause an aberrant splicing of the *hif-1* mRNA, removing 135 base pairs from the last exon (Figure 4A). We confirmed this

change in *hif-1* splicing by RT-PCR (Figure 4B). The aberrantly spliced *hif-1(nu469)* transcript is likely to have reduced stability.

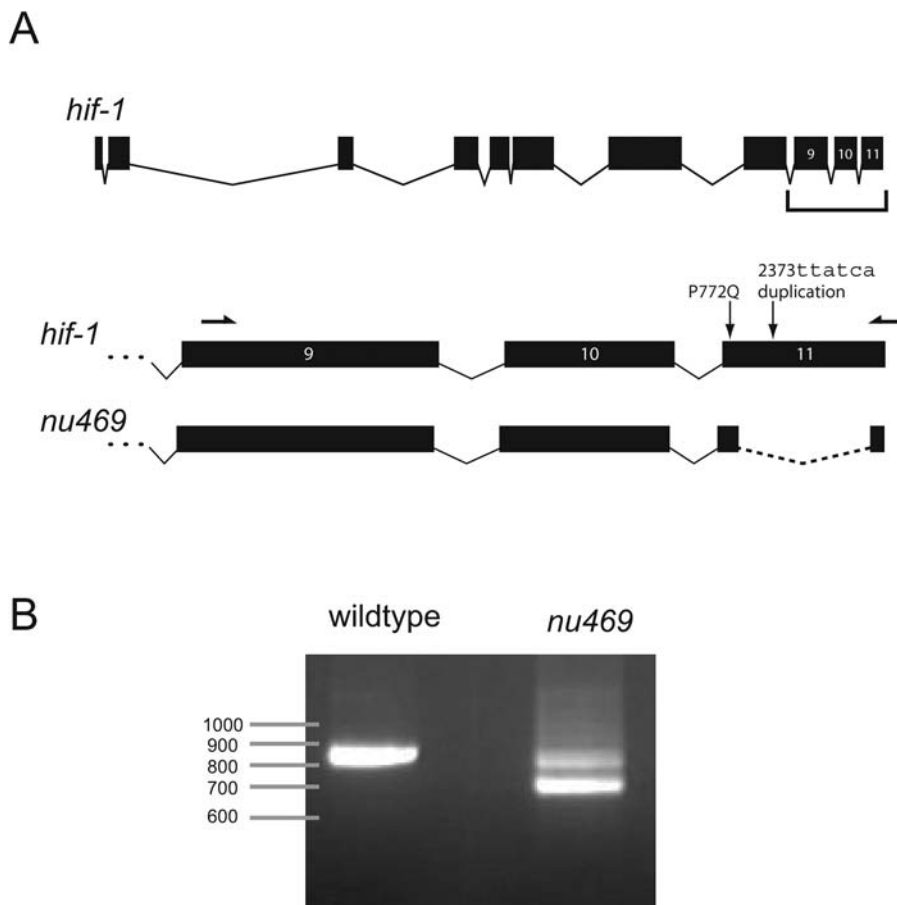
In aerobic conditions, the HIF-1 protein is constitutively degraded by the von Hippel-Lindau ubiquitin ligase [36–38]; consequently, the *hif-1(nu469)* mutation would presumably be phenotypically silent in normal growth conditions. The *hif-1(nu469)* mutation was not present in several other strains containing the *unc-43(n1186)* allele, suggesting that this mutation occurred spontaneously during culturing in our laboratory (data not shown). In summary, KP3365 animals carry a previously uncharacterized polymorphism in *hif-1*, which we identified based solely on our microarray hybridization results. Identifying such polymorphisms may allow researchers to explain unexpected aspects of mutant phenotypes of particular strains.

#### Using Microarrays to Identify a Mutation in Tomosyn, an Inhibitor of Neurotransmitter Secretion

To further address whether microarray hybridizations can be used to identify uncharacterized mutations, we analyzed a behavioral mutant that was isolated in a forward genetic screen for inhibitors of neurotransmitter secretion. Neurotransmission serves as the primary mode of communication between cells in the nervous system. Neurotransmitters such as acetylcholine (ACh) are secreted by presynaptic nerve cells, and activate receptors on postsynaptic cells. Behavioral and pharmacological screens in *C. elegans* have proven to be a powerful approach to identifying molecules involved in synaptic transmission and nervous system function [39–42]. The cholinesterase inhibitor aldicarb is widely used as a means to monitor ACh secretion at the *C. elegans* neuromuscular junction [41,43–46]. In the presence of aldicarb, ACh accumulates in the synaptic cleft, causing the body wall muscles to become hypercontracted and animals to become paralyzed. Mutations that increase ACh secretion cause hypersensitivity to the paralytic effects of aldicarb [46–48]. To identify negative regulators of ACh secretion, we used hypersensitivity to aldicarb as the basis for a forward genetic screen. One of the strongest mutations recovered in our screen was *nu468* (filled squares in Figure 5).

We meiotically mapped *nu468* to Chromosome 1, which contains approximately 2,700 genes. We then hybridized RNA from KP3293 *nu468* animals to the *C. elegans* GeneChip, comparing the hybridizations to wild-type hybridizations as previously described (Dataset S1). Six probesets showed significantly decreased expression in KP3293 animals (fold-change  $< -0.5$ ,  $p < 0.01$ ) (Figure 6), of which two corresponded to genes on Chromosome 1. Sequencing DNA from the mutant revealed a nonsense mutation in one of these genes, *tom-1*, the *C. elegans* ortholog of mammalian tomosyn (Figures 7A and S2). This lesion, a predicted NMD target, is consistent with the decreased transcript levels that we observed by the microarray hybridization. This is the only probe corresponding to *tom-1* on the *C. elegans* GeneChip.

We performed several experiments to confirm that the *tom-1(nu468)* mutation caused the aldicarb hypersensitivity observed in the KP3293 strain. First, we tested a second *tom-1* allele, *ok285*, which was generated by the *C. elegans* Gene Knockout Consortium (<http://celeganskoconsortium.omrf.org>). This allele, *tom-1(ok285)*, encodes a mutant protein lacking 202 residues in a highly conserved region, and



**Figure 4.** Characterization of the Splicing Defect in *hif-1(nu469)*

(A) Diagram of *hif-1* gene structure and the two mutations in *hif-1(nu469)*, a previously uncharacterized lesion in the background of the KP3365 strain. This lesion consists of two closely linked mutations: (1) C→A at position 2315 of the coding sequence, resulting in a P771Q mutation, and (2) an insertion of TTATCA after position 2373. Sequencing cDNA from KP3365 revealed that these two mutations result in the inappropriate splicing of the *hif-1* transcript (indicated by a dashed line), removing 135 base pairs of the last exon. Location of primers for PCR are indicated by half-arrows above exons 9 and 11.

(B) RT-PCR confirmed the altered splicing of the *hif-1(nu469)* mRNA.

DOI: 10.1371/journal.pgen.0010002.g004

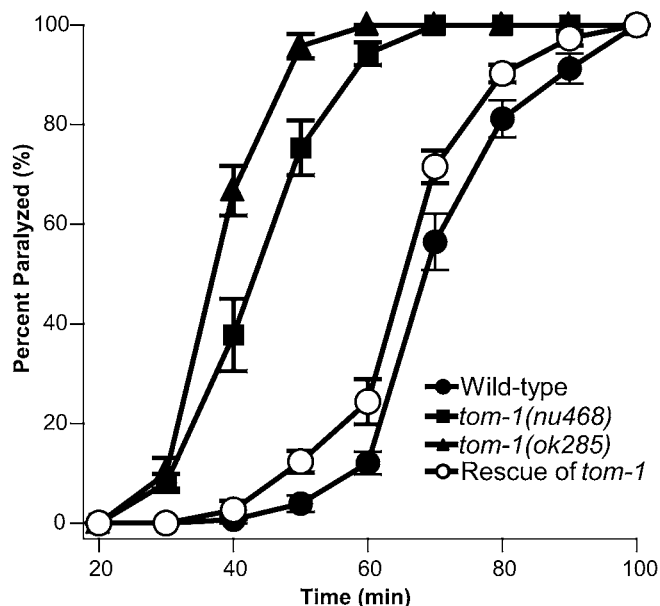
homozygous *tom-1(ok285)* mutants exhibited aldicarb hypersensitivity similar to that observed in KP3293 *tom-1(nu468)* mutants (triangles in Figure 5). Second, the aldicarb hypersensitivity of KP3293 *tom-1(nu468)* animals was rescued by a transgene driving expression of a *tom-1* cDNA in cholinergic neurons (open circles in Figure 5). Third, we found that a 4.2-kb *tom-1* promoter fragment was sufficient to drive expression of green fluorescent protein in the ventral cord motor neurons (Figure 7B), consistent with the idea that tomosyn acts as an inhibitor of ACh secretion in motor neurons. Tomosyn also showed expression in several neurons in head (Figure 7C) and tail ganglia (Figure 7D).

Tomosyn is an approximately 1,100-amino acid protein with two functional domains: (1) the C-terminal coiled-coil domain, which shares homology with synaptobrevin and has been shown to bind to syntaxin and SNAP-25, and (2) the approximately 600-residue WD40-rich N-terminal region, which shows strong homology to the fly tumor suppressor protein Lethal giant larvae (Figure 7A) [49,50]. Previous studies have shown that the synaptobrevin-like coiled-coil domain of rat tomosyn binds syntaxin and SNAP-25, forming a SNARE-like complex that occludes synaptobrevin [51]. This

suggests a mechanism whereby tomosyn competitively inhibits secretion by preventing SNARE complex formation. Supporting this hypothesis, overexpression of tomosyn in neuroendocrine cells results in a decrease in exocytosis in response to depolarization [49–52]. While these overexpression studies show that tomosyn can function to inhibit dense core vesicle release, they do not address the endogenous function of tomosyn. Our results provide the first in vivo evidence suggesting that endogenously expressed tomosyn inhibits neurotransmitter secretion in neurons.

### Generalizability of Microarray-Assisted Cloning

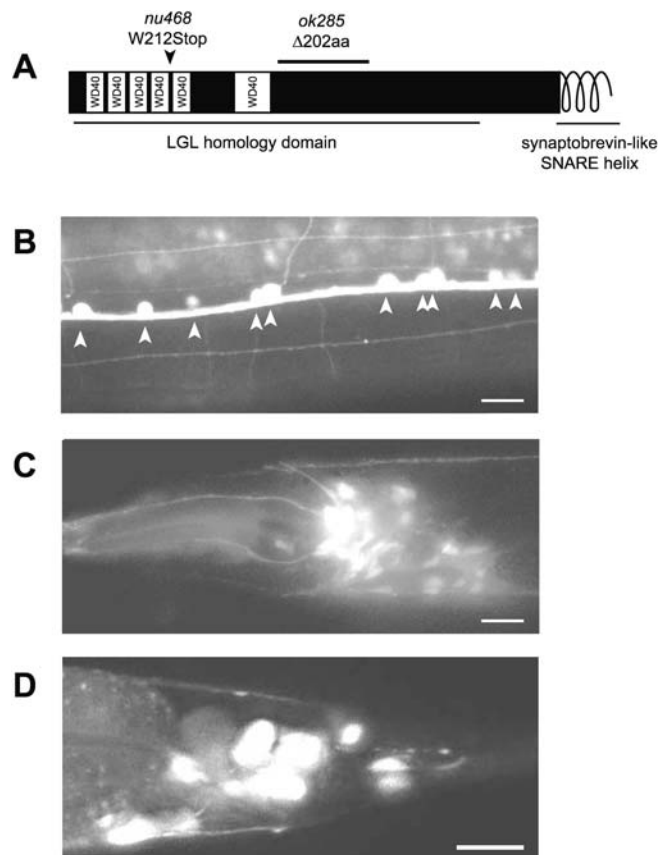
Since NMD functions in all eukaryotes [16,18], we wondered whether our strategy could be applied to other model systems. To address this, we conducted a retrospective analysis of microarray data from mutants in other organisms. We searched the public microarray databases for experiments in which researchers had analyzed mutants in other organisms. Specifically, we looked for hybridizations where mutant RNA had been compared to wild-type RNA and where the mutation was the result of a premature stop codon (and thus a predicted NMD target). For practical reasons, we also



**Figure 5.** Aldicarb Sensitivity and Rescue of KP3293 *nu468*

Levels of ACh secretion were assayed by following the time course of paralysis of animals on 1 mM aldicarb. Filled circle, wild-type; filled square, *nu468*, mutant recovered in our screen; filled triangle, *ok285* deletion allele of *tom-1*; open circle, *nu468* with a transgene expressing *tom-1* cDNA under the *unc-17* promoter. Data shown are averages from seven trials, except for *ok285*, which is an average of five trials. Error bars represent standard error.

DOI: 10.1371/journal.pgen.0010002.g005

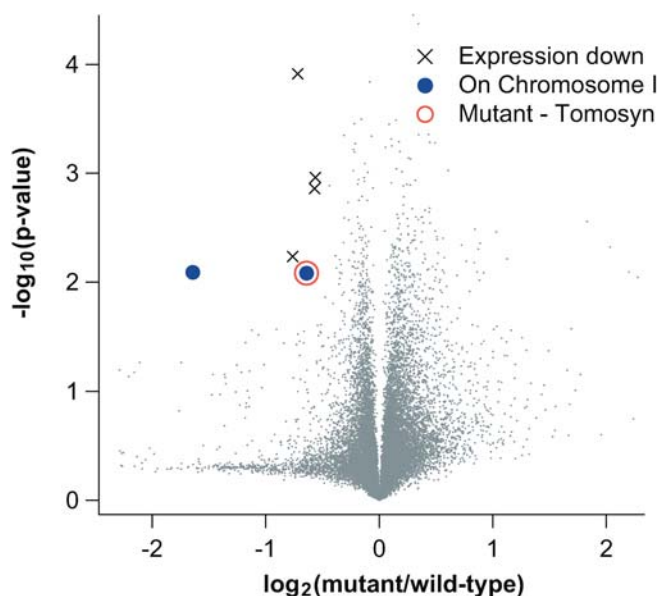


**Figure 7.** Expression of TOM-1, the *C. elegans* Ortholog of Tomosyn

(A) Schematic of worm tomosyn indicating the location of the premature stop found in *nu468* and deletion in *ok285*.

(B–D) Expression pattern of *tom-1* characterized with 4.2 kb of sequence upstream of the start codon driving expression of green fluorescent protein. Expression is seen in ventral cord motor neurons, with cell bodies indicated by arrowheads (B) and a number of neurons in the head (C) and the tail (D). Scale bars = 10  $\mu$ m.

DOI: 10.1371/journal.pgen.0010002.g007



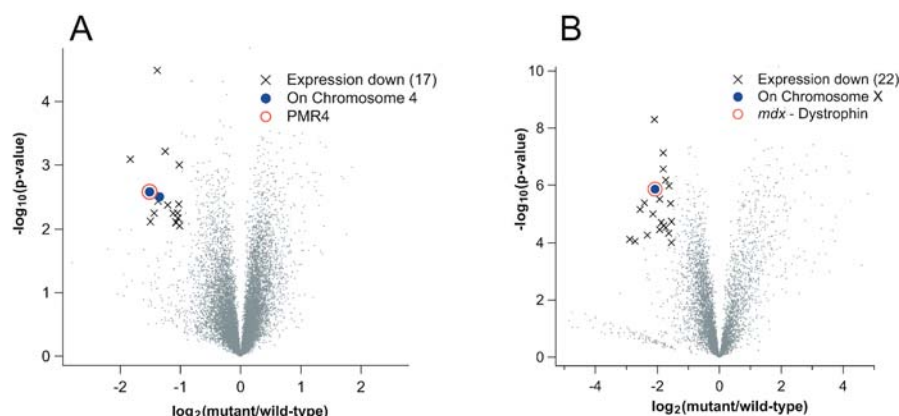
**Figure 6.** Positional Cloning of *tom-1(nu468)*

Expression data are illustrated as described in Figure 2. The symbol  $\times$  indicates probesets with reduced expression in KP3293 *nu468* (fold-change  $< -0.5$ ,  $p < 0.01$ ). Filled blue circles indicate probesets with reduced expression in KP3293 *nu468* that are also on Chromosome I. The open red circle indicates the probeset corresponding to *tom-1*. Sequencing of the *tom-1* gene in KP3293 *nu468* revealed a W212Stop mutation in the *tom-1* gene (see Figure 7A).

DOI: 10.1371/journal.pgen.0010002.g006

required that the mutant gene be represented and detectable on the microarray. Surprisingly, we found that only two experiments met these criteria. The first was a study of *pmr4* (*powdery mildew resistant 4*), a cell-wall biosynthesis gene in *Arabidopsis* that confers pathogen resistance when mutated. The lesion used in the microarray studies was a premature stop codon in the second exon (PMR4 dataset) [53]. The second was a study of the *mdx* mouse, an animal model of Duchenne muscular dystrophy, with a premature stop codon in exon 23 of dystrophin (MDX dataset) [54,55]. In both of these studies, the authors knew the nature of the mutation and were attempting to find genes whose expression changed in the mutant background.

For these two examples, we asked retrospectively whether hybridization data would have aided identification of the mutant genes. To do this, we reanalyzed the PMR4 and MDX data as described for the *C. elegans* mutants and constructed volcano plots (Figure 8). Each of these datasets had a different distribution on the fold-change and significance axes because of differences in the sample preparation, labeling efficiency, type of GeneChip, and number of hybridizations. We therefore adjusted our significance and ratio thresholds for differential expression (see Materials and Methods). In the



**Figure 8.** Analysis of Nonsense Mutants in *Arabidopsis* and Mouse

(A) PMR4 mutant; (B) MDX mutant. Expression data are illustrated as described in Figure 2. The symbol  $\times$  indicates probesets with reduced expression in the nonsense mutant (fold-change  $< -1.0$ ,  $p < 0.01$ ) for PMR4 and (fold-change  $< -1.5$ ,  $p < 0.0001$ ) for MDX. Numbers of genes with significantly reduced expression are indicated for both mutants. Filled blue circles indicate probesets with reduced expression that are on same chromosome as the mutant gene. The open red circle indicates the probeset corresponding to the mutant gene.

DOI: 10.1371/journal.pgen.0010002.g008

PMR4 hybridizations, 17 genes showed significantly decreased expression (fold-change  $< -1.0$ ,  $p < 0.01$ ). Of these, two genes were on Chromosome 4, the most significant of which was PMR4 (Figure 8A). In the case of the MDX data, 22 genes showed significantly decreased expression (fold-change  $< -1.5$ ,  $p < 0.0001$ ). Of these, only the dystrophin gene was on X (Figure 8B). In each case, there was only one probe on the GeneChip corresponding to the mutant gene. For both of these examples, the combination of microarray data and chromosomal mapping quickly reduced the number of candidates to one or two genes.

## Discussion

We present evidence demonstrating the utility of microarray hybridizations in facilitating the rapid identification of mutations isolated in forward genetic screens. Several results suggest that this technique will be widely applicable. This strategy was successful in identification of *C. elegans*, mouse, and *Arabidopsis* mutations. Mutations were successfully identified in both transcription factors and signal transduction components, which are likely to be the most challenging cases. Mutant genes were successfully detected using data obtained with both printed arrays and Affymetrix chips. And finally, we were able to identify two previously uncharacterized *C. elegans* mutations with this approach.

Will this strategy work for genes that regulate the expression of many other genes? We provide examples for successful identification of three genes that *directly* affect transcription—two transcription factors (*mec-3* and *hif-1*) and a protein kinase that regulates neuronal gene expression (*unc-43*). Although 70 genes were differentially expressed in *mec-3* mutants, only three differentially expressed genes mapped within a 2-cM interval containing *mec-3* and 100 other genes. Therefore, microarray hybridizations would have facilitated identification of *mec-3*.

The success rate for this strategy depends on three factors: (1) the fraction of genes that are detectable by microarray, (2) the frequency of nonsense alleles recovered in screens, and (3) the efficiency with which nonsense mutated mRNAs are degraded by NMD. In our hybridizations using mRNA

prepared from whole worms, 80% of the genes on the array showed detectable expression. In cases where a mutation affects a particular cell type or tissue, the likelihood of detecting a particular transcript can be increased using RNA isolated from that tissue or cell type [28,56].

What fraction of newly isolated mutations will be nonsense alleles (see Figure 1)? Our analysis suggests that for 20% of *C. elegans* genes, nonsense mutations are rarely recovered. For the remaining 80% of *C. elegans* genes, 45% of alleles recovered were nonsense alleles. Given these ratios, the rate of successful gene identification by microarray hybridization could be increased by analyzing multiple alleles of the same gene and by selecting genes for which null alleles are the most frequently isolated class of alleles recovered in screens. Furthermore, microarray-assisted cloning will likely be useful for other categories of alleles that decrease transcript abundance, e.g., the spontaneous *hif-1(mu469)* mutation. These include mutations altering pre-mRNA splicing, mutations in promoters, frameshift mutations, and mutations yielding transcripts that lack termination codons [57–59].

What fraction of nonsense alleles are efficiently targeted by the NMD machinery? In each of the six examples we present above, this was the case, but how often do nonsense transcripts evade degradation by the NMD machinery? Rules governing NMD recognition of mutant mRNAs have been described in yeast, *C. elegans*, and mammals [16–18,59–63]. The NMD machinery distinguishes premature stop codons from natural stops using the exon-junction complexes that are deposited at exon–exon boundaries by the spliceosome. Stops that are greater than 50–55 base pairs upstream of the last exon-junction complex are recognized by the NMD machinery as premature and are efficiently targeted for destruction [61,64]. Prior studies have shown that 100% ( $n = 23$ ) of *C. elegans* nonsense mutations were susceptible to NMD surveillance (measured either by mRNA abundance or by suppression of mutant phenotypes by NMD pathway mutations) [17]. Of these, six mutations (26%) were judged to be only partially targeted by NMD. Based on these examples and those we describe here, we estimate that 75%–100% of nonsense alleles in *C. elegans* would show a detectable decrease in mRNA



levels. Considering all three of these factors (gene detection by microarray, nonsense allele frequency, and NMD efficiency), we expect microarray-assisted cloning to be successful in 25%–30% of positional clonings (assuming only one allele is hybridized per gene).

The principal costs of positional cloning are those incurred in isolating, phenotyping, and genotyping a sufficient number of recombinants (i.e., informative meioses) to map a mutation to a small genetic interval. A typical positional cloning requires 2,000–10,000 informative meioses. Our results suggest that microarray hybridizations can significantly reduce the number of meioses required for positional clonings. In five of six cases, microarray data in conjunction with chromosomal linkage data were sufficient for gene identification. Therefore, while we expect that this strategy will be useful in many genetic systems, microarray-assisted cloning promises to provide the greatest value in organisms such as mouse and zebrafish, where long generation times and large genomes make meiotic mapping more time consuming and costly. Furthermore, microarray-assisted cloning may be particularly useful in cases where mutant phenotypes are more difficult to assess, such as behavioral mutants, or incompletely penetrant or complex (i.e., multigenic) traits [65,66]. Given the effort and challenges involved in meiotic mapping and the relative ease and speed of microarray hybridizations, we believe that this microarray-based strategy provides significant benefit, even though it will be successful in only a subset of cases.

Can microarrays be used to aid the cloning of human disease genes? One-third of human disease genes are predicted to be caused by nonsense lesions or mutations that decrease transcript abundance [21,22]. Furthermore, nonsense mutant transcripts encoded by disease genes such as *BRCA1* and *hepatocyte nuclear factor 1 $\alpha$*  have been shown to be effectively degraded by NMD [67,68]. Given the enormous time and expense involved in mapping genes in humans, the strategy described here could provide a valuable addition to the toolbox of human geneticists.

## Materials and Methods

**Allele analysis.** Information about 930 recessive single base-pair substitution alleles was downloaded from WormBase (<http://www.wormbase.com>), Release WS123 (see Table S1). Information about 82 additional alleles was obtained through literature searches. Based on their molecular description, 943 alleles were classified as either NMD targets (nonsense) or non-NMD targets (missense). Excluded from the analysis were 69 alleles that could not be definitively classified. These alleles included those with incomplete molecular descriptions and those with lesions such as splice site mutations that could not be classified without further characterization.

**RNA sample preparation.** Animals analyzed in microarray experiments were first synchronized by hypochlorite treatment and arrested at the first larval stage by incubation for 22 h in M9 [27,69]. Animals were then grown at 20 °C on 15-cm NG HB101 plates until the fourth larval stage (approximately 46 h). Animals were washed, harvested in M9, and then flash-frozen in liquid nitrogen and stored at –80 °C. Total RNA was prepared by Trizol extraction (Invitrogen, Carlsbad, California, United States).

**Microarray target preparation and hybridization.** Targets were prepared and hybridized at the Harvard Medical School Biopolymer Facility. Starting with 10  $\mu$ g of total RNA, first-strand cDNA was synthesized as described in the Affymetrix (Santa Clara, California, United States) expression technical manual. Briefly, 10  $\mu$ g of RNA was added to 1  $\mu$ l of 50  $\mu$ M T7 primer (HPLC purified) (Integrated DNA Technologies, Coralville, Iowa, United States) in a volume of 9  $\mu$ l. Then 1  $\mu$ l of each Poly A spike control (5 nM) was added to the RNA,

and T7 was added as an internal control. Poly A spikes were created from Poly A-tailed genes from *Bacillus subtilis* cloned into Stratagene (La Jolla, California, United States) pBluescript as an XhoI-to-NotI insert 5'–3', respectively, and commercially available through ATCC (Manassas, Virginia, United States) (see Affymetrix technical expression manual). The RNA, T7, and Poly A spike controls were heated to 70 °C for 10 min and then placed on ice for 5 min. The RNA, T7, and Poly A mix was then heated to 42 °C. Then 4  $\mu$ l of 5 $\times$  first-strand buffer (Invitrogen), 2  $\mu$ l of 0.1 M DTT (Invitrogen), 1  $\mu$ l of 10 mM dNTP (Invitrogen), and 1  $\mu$ l of Superscript II, RNase H– was added to the RNA and incubated at 42 °C for 1 h. Double-strand DNA was created via a replacement reaction under the following conditions. To the 20- $\mu$ l first-strand reaction was added 91  $\mu$ l of water, 30  $\mu$ l of second-strand buffer (Invitrogen), 3  $\mu$ l of 10 mM dNTP (Invitrogen), 1  $\mu$ l of *Escherichia coli* DNA ligase (Invitrogen), 1  $\mu$ l of RNase H (Invitrogen), and 4  $\mu$ l of *E. coli* DNA polymerase (Invitrogen). This 130- $\mu$ l second-strand mix was added to the first-strand reaction and incubated at 16 °C for 2 h, then 2  $\mu$ l of T4 DNA polymerase was added for 5 min at 16 °C, then the reaction was phenol-chloroform-extracted using 150  $\mu$ l of phenol chloroform isoamyl alcohol (pH 7) (Ambion, Austin, Texas, United States), and the organic and aqueous phases were separated using a 1.5-ml phase lock heavy gel (Brinkmann Eppendorf, Westbury, New York, United States). The 150- $\mu$ l aqueous layer was removed and precipitated in 375  $\mu$ l of 100% ethanol and 15  $\mu$ l of 3 M sodium acetate (Sigma, St. Louis, Missouri, United States). The cDNA pellet was isolated using an Eppendorf (Hamburg, Germany) 5415C centrifuge at room temperature for 20 min. Ethanol was aspirated and the pellet washed in 75% ethanol, centrifuged for 10 min, and aspirated. The cDNA pellet was rehydrated using 22  $\mu$ l of nuclease-free water (Ambion) and used with the BioArray HighYield RNA Transcript Labeling Kit T7 (Enzo Life Sciences, Farmingdale, New York, United States). The resulting biotinylated cRNA probes were purified using RNeasy columns (Qiagen, Valencia, California, United States) and quantitated using A260 with an Agilent (Palo Alto, California, United States) 8453 spectrophotometer. Then 15  $\mu$ g of labeled probe was fragmented with 5 $\times$  fragmentation buffer (see Affymetrix technical manual) and combined with hybridization controls (Affymetrix), herring sperm DNA (Promega, Madison, Wisconsin, United States), and BSA (Invitrogen) to create 300  $\mu$ l of hybridization mix. Of this, 200  $\mu$ l was added to the Affymetrix *C. elegans* GeneChip. Hybridization was done in a GeneChip Hybridization Oven 320 for 16 h at 45 °C, processed on an Affymetrix Fluidics Station 400 using double amplification staining (see Affymetrix technical manual), and washed using fluidics protocol EukGE-WS2v4. The GeneChips were then scanned on a Hewlett-Packard (Palo Alto, California, United States) GeneArray Scanner.

**Public datasets.** Descriptions of all available hybridizations in the *C. elegans* whole-genome expression profiles were downloaded from the Stanford Microarray Database (<http://genome-www5.stanford.edu>). These were then searched to find direct mutant versus wild-type comparisons. The 368 hybridizations that were publicly accessible included a large number of developmental time courses, aging experiments, and heat-shock and tissue-specific expression profiles (see Figure 2 in [27] for more detail). The only direct mutant versus wild-type comparison was the *mec-3(e1338)* analysis, which consisted of six hybridizations. For these experiments, normalized log expression ratios were downloaded from the Stanford Microarray Database (ExptSetNo = 1461). Affymetrix expression values for mouse and *Arabidopsis* datasets were downloaded from NCBI's Gene Expression Omnibus (GEO, <http://www.ncbi.nlm.nih.gov/geo>).

**Microarray data analysis.** For Affymetrix data, probesets were first filtered to eliminate those that showed no detectable signal. A threshold of 32 was used for the *C. elegans* and *Arabidopsis* data. A threshold of 256 was used for the *mdx* data because these data showed significantly higher signals than the other datasets. This is most likely because the RNA for these experiments was prepared from a single tissue (mouse skeletal muscle), as opposed to the *C. elegans* and *Arabidopsis* RNA, which was derived from the whole organism. For printed arrays, only spots that showed detectable signal (mean signals greater than 1.5 standard deviations above background) were included in the analysis.

Probes were classified as differentially expressed based on two criteria: fold-change and statistical significance. For Affymetrix data, fold-change was calculated as the average expression in the mutant divided by average expression in wild-type. For printed arrays (*mec-3*), fold-change was calculated by averaging the expression ratio in each of the mutant-versus-wild-type replicate hybridizations. This ratio provides a measure of the magnitude of expression difference between mutant and wild-type samples. To assess the statistical

significance of expression differences, we compared the replicate expression values in the mutant hybridizations to replicate expression values in the wild-type hybridizations using a Student's *t*-test, and calculated a *p*-value. We then constructed a volcano plot with the  $\log_2(\text{fold-change})$  on the x-axis and negative  $\log_{10}(p\text{-value})$  on the y-axis [31]. Cutoffs for differential expression were based on shape and distribution of individual volcano plots.

Raw image files were converted to probeset data (.cel files) in Microarray Suite (MAS 5.0). The nine probeset data files were normalized together and expression values were determined using the Robust Multi-chip Average (RMA) method as implemented in RMA Express (<http://stat-www.berkeley.edu/~bolstad/RMAExpress/RMAExpress.html>). Subsequent analysis was done using the R statistical computing package (<http://www.r-project.org>) and the Bioconductor libraries (<http://www.bioconductor.org>). Graphs were produced in Igor Pro 4.0 (WaveMetrics, Lake Oswego, Oregon, United States). Probeset annotations were downloaded from the Affymetrix Web site (<http://www.affymetrix.com>).

**Molecular characterization of *hif-1(nu469)*.** Using RNA prepared from KP3365 and wild-type animals, first-strand cDNA was synthesized using a primer specific for the 3' UTR of *hif-1* (Invitrogen). The *hif-1* gene was then amplified by PCR from this cDNA and sequenced.

**Isolation and mapping of the *tom-1* mutation.** The *tom-1(nu468)* allele was isolated in an ethyl methane sulfonate screen for mutants that displayed hypersensitivity to aldicarb. F<sub>2</sub> progeny of mutagenized animal were transferred to agar plates containing 0.5 mM aldicarb (Chem Service, West Chester, Pennsylvania, United States). After 1 h, a time point at which all wild-type worms were still moving, paralyzed animals were transferred to separate plates and rescreened for aldicarb sensitivity in subsequent generations. *nu468* was determined to be recessive and was mapped to Chromosome 1 using conventional meiotic mapping.

**Analysis of aldicarb sensitivity.** Aldicarb sensitivity was assessed essentially as described [45,46]. Briefly, for each experiment, 20 to 25 animals were transferred to agar plates containing 1 mM aldicarb (Chem Service). Paralysis was assessed every 10 min by prodding each animal with a platinum wire. Data from independent trials were averaged and used to calculate standard error. All experiments were conducted blind with respect to the genotype of the animals.

**Molecular characterization of *tom-1(ok285)*.** Genomic sequence from VC223 animals, available on WormBase, shows a deletion of 1,580 nucleotides that removes all of exons 11–13 and part of exon 10. To characterize the effect of this lesion on the *tom-1* gene product, we purified RNA from VC223 animals and amplified the mutated *tom-1* mRNA by RT-PCR (Invitrogen). Sequencing revealed that the genomic deletion results in an in-frame lesion in the mRNA, removing 606 nucleotides of coding sequence, and adding 23 nucleotides of intronic sequence and a 490-base-pair alternative exon from isoform C of *tom-1* that is located just downstream of the deletion.

**Rescue of *tom-1(nu468)*.** *tom-1(nu468)* was rescued using the full-length cDNA of the major splice form of *tom-1* (M01A10.2a) under the promoter of *unc-17* synaptic vesicle ACh transporter. A transgenic strain was isolated by microinjecting the rescuing plasmid at 100 ng/μl using *ptt-3::dsRed* as a marker into *nu468*. During characterization and rescue of *tom-1*, we discovered that the start site and first exon were incorrectly predicted and described in WormBase. We identified the correct start site and initial two exons of *tom-1* by performing RT-PCR using a primer complementary to the trans-splice acceptor (SL1). This corrected version of *tom-1* shows much better alignment to the N-terminus of mammalian homologs (see Figure S2).

## Supporting Information

**Dataset S1.** Microarray Expression Data from Wild-Type, *unc-43(n1186)*, and *tom-1(nu468)* Animals

Found at DOI: 10.1371/journal.pgen.0010002.sd001 (2 MB TDS).

**Figure S1.** Probeset Alignments to *unc-43* CaMKII Isoform H (K11E8.1h)

Target sequences were downloaded (<http://www.affymetrix.com>) and aligned to *unc-43* CaMKII isoform H. Since a majority of genes on the *C. elegans* GeneChip are represented by one probeset, *unc-43* represents an atypical case. To explain this, it is useful to consider the history and design of the *C. elegans* GeneChip (see [http://www.affymetrix.com/support/technical/datasheets/celegans\\_drosophila\\_datasheet.pdf](http://www.affymetrix.com/support/technical/datasheets/celegans_drosophila_datasheet.pdf)). The targets for this GeneChip were designed by

Affymetrix based on more than 18,800 Sanger Center predicted transcripts from the December 2000 genome sequence, as well as 2,300 3' EST clusters and 300 GenBank mRNAs (Release 121). Despite efforts to eliminate redundancy, there is not a strict one-to-one correspondence between the current set of genes and the probesets on the GeneChip. Our analysis indicates that 13% of the genes on the GeneChip are represented by more than one probe. Even so, having *unc-43* CaMKII represented by seven probesets is an unusual situation. However, only two of these probesets (193459\_s\_at and 193463\_s\_at) are annotated as corresponding to the *unc-43* CaMKII mRNA transcript according to annotations of the *C. elegans* GeneChip provided by Affymetrix in the March 28, 2003, update (downloadable at <http://www.affymetrix.com>). During our examination of the eight candidate probesets that showed decreased expression and were on Chromosome 4, we discovered five additional probesets that corresponded to *unc-43* CaMKII. Four of these probesets (172058\_x\_at, 175820\_s\_at, 175821\_s\_at, and 175824\_s\_at) were based on GenBank sequences, and one (187759\_s\_at) was based on a predicted open reading frame, Y43C5B.1, that was part of the genome as of December 2000 but has since been shown to correspond to the 5' UTR of *unc-43* CaMKII. These additional probesets provided a serendipitous blind control, since we were not aware of their existence until they appeared on our candidate list from the hybridization. There is also an additional (eighth) probeset (173423\_at) described in the Affymetrix annotation as corresponding to *unc-43* CaMKII. However, based on the current gene model, 173423\_at aligns to the intron between exons 11 and 12. As would be expected, this probeset shows no detectable expression and thus was not considered in our analysis of the *unc-43* CaMKII mutant.

Found at DOI: 10.1371/journal.pgen.0010002.sg001 (33 KB PDF).

**Figure S2.** Alignment of TOM-1 to Human and Rat Tomosyn

Sequences of human and mouse tomosyn were downloaded from Ensembl (<http://www.ensembl.org>). The multiple sequence alignment was performed using T-Coffee (<http://www.ch.embnet.org/software/TCoffee.html>). Alignment output was produced using GeneDoc (<http://www.psc.edu/biomed/genedoc>).

Found at DOI: 10.1371/journal.pgen.0010002.sg002 (104 KB PDF).

**Table S1.** List of Missense and Nonsense Alleles

Found at DOI: 10.1371/journal.pgen.0010002.st001 (51 KB PDF).

## Accession Numbers

The National Center for Biotechnology Information Gene Expression Omnibus (GEO, <http://www.ncbi.nlm.nih.gov/geo>) accession number for the data generated by the authors and discussed in this publication is GSE2210. The GEO accession numbers for datasets downloaded by the authors and discussed in this paper are *mdx* mouse (GDS236), *mdx* mutant (GDS236), PMR4 *Arabidopsis* (GDS417), and PMR4 mutant (GDS417).

The GenBank (<http://www.ncbi.nlm.nih.gov/Genbank>) accession numbers for genes and gene products discussed in this paper are *tom-1(ok285)* (AY912103) and tomosyn (AY912102).

The Ensembl (<http://www.ensembl.org/>) accession numbers for genes and gene products discussed in this paper are human tomosyn (ENSP00000179882) and mouse tomosyn (ENSRNOP00000018806).

## Acknowledgments

We thank K. Vranizan for statistical and programming advice; T. Rector for Affymetrix sample preparation and hybridizations; J. Rine, D. Altshuler, V. Mootha, S. McCarroll, and C. Murphy for helpful discussions; T. Speed for statistical input; P. Juo, L. Drier, J. Dittman, and I. Ruvinsky for reading this manuscript; other members of the Kaplan and Ruvkun labs for suggestions; K. Scott for providing a fantastic foster lab for MD in Berkeley; and Gregoire for sustenance. Some of the strains described here were provided by the *C. elegans* Genetic Stock Center. This work was supported by a grant from the National Institutes of Health to JK (GM54728). MD was supported by a Howard Hughes Medical Institute predoctoral fellowship.

**Competing interests.** The authors have declared that no competing interests exist.

**Author contributions.** MD, JN, and JMK conceived and designed the experiments. MD performed the experiments and analyzed the data. MD and JMK wrote the paper. ■

## References

- Haffter P, Granato M, Brand M, Mullins MC, Hammerschmidt M, et al. (1996) The identification of genes with unique and essential functions in the development of the zebrafish, *Danio rerio*. *Development* 123: 1–36.
- Driever W, Solnica-Krezel L, Schier AF, Neuhauss SC, Malicki J, et al. (1996) A genetic screen for mutations affecting embryogenesis in zebrafish. *Development* 123: 37–46.
- Zarbalis K, May SR, Shen Y, Ekker M, Rubenstein JL, et al. (2004) A focused and efficient genetic screening strategy in the mouse: Identification of mutations that disrupt cortical development. *PLoS Biol* 2: DOI: 10.1371/journal.pbio.0020219
- O'Brien TP, Frankel WN (2004) Moving forward with chemical mutagenesis in the mouse. *J Physiol* 554: 13–21.
- Nadeau JH, Balling R, Barsh G, Beier D, Brown SD, et al. (2001) Sequence interpretation. Functional annotation of mouse genome sequences. *Science* 291: 1251–1255.
- Bessereau JL, Wright A, Williams DC, Schuske K, Davis MW, et al. (2001) Mobilization of a *Drosophila* transposon in the *Caenorhabditis elegans* germ line. *Nature* 413: 70–74.
- Cooley L, Kelley R, Spradling A (1988) Insertional mutagenesis of the *Drosophila* genome with single P elements. *Science* 239: 1121–1128.
- Gaiano N, Amsterdam A, Kawakami K, Allende M, Becker T, et al. (1996) Insertional mutagenesis and rapid cloning of essential genes in zebrafish. *Nature* 383: 829–832.
- Breindl M, Harbers K, Jaenisch R (1984) Retrovirus-induced lethal mutation in collagen I gene of mice is associated with an altered chromatin structure. *Cell* 38: 9–16.
- Korswagen HC, Durbin RM, Smits MT, Plasterk RH (1996) Transposon Tc1-derived, sequence-tagged sites in *Caenorhabditis elegans* as markers for gene mapping. *Proc Natl Acad Sci U S A* 93: 14680–14685.
- Jansen G, Hazendonk E, Thijssen KL, Plasterk RH (1997) Reverse genetics by chemical mutagenesis in *Caenorhabditis elegans*. *Nat Genet* 17: 119–121.
- Rong YS, Golik KG (2000) Gene targeting by homologous recombination in *Drosophila*. *Science* 288: 2013–2018.
- Bradley A, Zheng B, Liu P (1998) Thirteen years of manipulating the mouse genome: A personal history. *Int J Dev Biol* 42: 943–950.
- Stemple DL (2004) TILLING—A high-throughput harvest for functional genomics. *Nat Rev Genet* 5: 145–150.
- McCallum CM, Comai L, Greene EA, Henikoff S (2000) Targeting induced local lesions in genomes (TILLING) for plant functional genomics. *Plant Physiol* 123: 439–442.
- Hentze MW, Kulozik AE (1999) A perfect message: RNA surveillance and nonsense-mediated decay. *Cell* 96: 307–310.
- Mango SE (2001) Stop making nonSense: The *C. elegans smg* genes. *Trends Genet* 17: 646–653.
- Wilusz CJ, Wang W, Peltz SW (2001) Curbing the nonsense: The activation and regulation of mRNA surveillance. *Genes Dev* 15: 2781–2785.
- Huuskio P, Ponciano-Jackson D, Wolf M, Kiefer JA, Azorsa DO, et al. (2004) Nonsense-mediated decay microarray analysis identifies mutations of *EPHB2* in human prostate cancer. *Nat Genet* 36: 979–983.
- Noensie EN, Dietz HC (2001) A strategy for disease gene identification through nonsense-mediated mRNA decay inhibition. *Nat Biotechnol* 19: 434–439.
- Frischmeyer PA, Dietz HC (1999) Nonsense-mediated mRNA decay in health and disease. *Hum Mol Genet* 8: 1893–1900.
- Mendell JT, Dietz HC (2001) When the message goes awry: Disease-producing mutations that influence mRNA content and performance. *Cell* 107: 411–414.
- Raser JM, O'Shea EK (2004) Control of stochasticity in eukaryotic gene expression. *Science* 304: 1811–1814.
- Fraser HB, Hirsh AE, Giaever G, Kumm J, Eisen MB (2004) Noise minimization in eukaryotic gene expression. *PLoS Biol* 2: DOI: 10.1371/journal.pbio.0020137
- Brem RB, Yvert G, Clinton R, Kruglyak L (2002) Genetic dissection of transcriptional regulation in budding yeast. *Science* 296: 752–755.
- Schadt EE, Monks SA, Drake TA, Lusis AJ, Che N, et al. (2003) Genetics of gene expression surveyed in maize, mouse and man. *Nature* 422: 297–302.
- Kim SK, Lund J, Kiraly M, Duke K, Jiang M, et al. (2001) A gene expression map for *Caenorhabditis elegans*. *Science* 293: 2087–2092.
- Zhang Y, Ma C, Delohery T, Nasipak B, Foat BC, et al. (2002) Identification of genes expressed in *C. elegans* touch receptor neurons. *Nature* 418: 331–335.
- Way JC, Chalfie M (1988) *mec-3*, a homeobox-containing gene that specifies differentiation of the touch receptor neurons in *C. elegans*. *Cell* 54: 5–16.
- Xue D, Tu Y, Chalfie M (1993) Cooperative interactions between the *Caenorhabditis elegans* homeoproteins UNC-86 and MEC-3. *Science* 261: 1324–1328.
- Cui X, Churchill GA (2003) Statistical tests for differential expression in cDNA microarray experiments. *Genome Biol* 4: 210.
- Way JC, Chalfie M (1989) The *mec-3* gene of *Caenorhabditis elegans* requires its own product for maintained expression and is expressed in three neuronal cell types. *Genes Dev* 3: 1823–1833.
- Reiner DJ, Newton EM, Tian H, Thomas JH (1999) Diverse behavioural defects caused by mutations in *Caenorhabditis elegans unc-43* CaM kinase II. *Nature* 402: 199–203.
- Sagasti A, Hisamoto N, Hyodo J, Tanaka-Hino M, Matsumoto K, et al. (2001) The CaMKII UNC-43 activates the MAPKKK NSY-1 to execute a lateral signaling decision required for asymmetric olfactory neuron fates. *Cell* 105: 221–232.
- Maxwell PH, Wiesener MS, Chang GW, Clifford SC, Vaux EC, et al. (1999) The tumour suppressor protein VHL targets hypoxia-inducible factors for oxygen-dependent proteolysis. *Nature* 399: 271–275.
- Epstein AC, Gleadle JM, McNeill LA, Hewitson KS, O'Rourke J, et al. (2001) *C. elegans* EGL-9 and mammalian homologs define a family of dioxygenases that regulate HIF by prolyl hydroxylation. *Cell* 107: 43–54.
- Jaakkola P, Mole DR, Tian YM, Wilson MI, Gielbert J, et al. (2001) Targeting of HIF- $\alpha$  to the von Hippel-Lindau ubiquitination complex by O<sub>2</sub>-regulated prolyl hydroxylation. *Science* 292: 468–472.
- Ivan M, Kondo K, Yang H, Kim W, Valiando J, et al. (2001) HIF $\alpha$  targeted for VHL-mediated destruction by proline hydroxylation: Implications for O<sub>2</sub> sensing. *Science* 292: 464–468.
- Trent C, Tsuing N, Horvitz HR (1983) Egg-laying defective mutants of the nematode *Caenorhabditis elegans*. *Genetics* 104: 619–647.
- Avery L (1993) The genetics of feeding in *Caenorhabditis elegans*. *Genetics* 133: 897–917.
- Miller KG, Alfonso A, Nguyen M, Crowell JA, Johnson CD, et al. (1996) A genetic selection for *Caenorhabditis elegans* synaptic transmission mutants. *Proc Natl Acad Sci U S A* 93: 12593–12598.
- Nonet ML (1999) Studying mutants that affect neurotransmitter release in *C. elegans*. In: Bellen H editor. *Neurotransmitter release*. New York: Oxford University Press. pp. 265–303.
- Nguyen M, Alfonso A, Johnson CD, Rand JB (1995) *Caenorhabditis elegans* mutants resistant to inhibitors of acetylcholinesterase. *Genetics* 140: 527–535.
- Nonet ML, Grundahl K, Meyer BJ, Rand JB (1993) Synaptic function is impaired but not eliminated in *C. elegans* mutants lacking synaptotagmin. *Cell* 73: 1291–1305.
- Lackner MR, Nurrish SJ, Kaplan JM (1999) Facilitation of synaptic transmission by EGL-30 G $\alpha$  and EGL-8 PLC $\beta$ : DAG binding to UNC-13 is required to stimulate acetylcholine release. *Neuron* 24: 335–346.
- Nurrish S, Segalat L, Kaplan JM (1999) Serotonin inhibition of synaptic transmission: G $\alpha$ (0) decreases the abundance of UNC-13 at release sites. *Neuron* 24: 231–242.
- van der Linden AM, Simmer F, Cuppen E, Plasterk RH (2001) The G-protein beta-subunit GPB-2 in *Caenorhabditis elegans* regulates the G $\alpha$ ( $\alpha$ )-G $\alpha$ ( $\alpha$ ) signaling network through interactions with the regulator of G-protein signaling proteins EGL-10 and EAT-16. *Genetics* 158: 221–235.
- Wang ZW, Saifee O, Nonet ML, Salkoff L (2001) SLO-1 potassium channels control quantal content of neurotransmitter release at the *C. elegans* neuromuscular junction. *Neuron* 32: 867–881.
- Masuda ES, Huang BC, Fisher JM, Luo Y, Scheller RH (1998) Tomosyn binds t-SNARE proteins via a VAMP-like coiled coil. *Neuron* 21: 479–480.
- Fujita Y, Shiratani H, Sakisaka T, Asakura T, Ohya T, et al. (1998) Tomosyn: A syntaxin-1-binding protein that forms a novel complex in the neurotransmitter release process. *Neuron* 20: 905–915.
- Hatsuzawa K, Lang T, Fasshauer D, Bruns D, Jahn R (2003) The R-SNARE motif of tomosyn forms SNARE core complexes with syntaxin 1 and SNAP-25 and down-regulates exocytosis. *J Biol Chem* 278: 31159–31166.
- Yizhar O, Matti U, Melamed R, Hagalili Y, Bruns D, et al. (2004) Tomosyn inhibits priming of large dense-core vesicles in a calcium-dependent manner. *Proc Natl Acad Sci U S A* 101: 2578–2583.
- Nishimura MT, Stein M, Hou BH, Vogel JP, Edwards H, et al. (2003) Loss of a callose synthase results in salicylic acid-dependent disease resistance. *Science* 301: 969–972.
- Sicinski P, Geng Y, Ryder-Cook AS, Barnard EA, Darlison MG, et al. (1989) The molecular basis of muscular dystrophy in the mdx mouse: A point mutation. *Science* 244: 1578–1580.
- Tseng BS, Zhao P, Pattison JS, Gordon SE, Granchelli JA, et al. (2002) Regenerated mdx mouse skeletal muscle shows differential mRNA expression. *J Appl Physiol* 93: 537–545.
- Roy PJ, Stuart JM, Lund J, Kim SK (2002) Chromosomal clustering of muscle-expressed genes in *Caenorhabditis elegans*. *Nature* 418: 975–979.
- Frischmeyer PA, van Hoof A, O'Donnell K, Guerrero AL, Parker R, et al. (2002) An mRNA surveillance mechanism that eliminates transcripts lacking termination codons. *Science* 295: 2258–2261.
- van Hoof A, Frischmeyer PA, Dietz HC, Parker R (2002) Exosome-mediated recognition and degradation of mRNAs lacking a termination codon. *Science* 295: 2262–2264.
- Wickens M, Anderson P, Jackson RJ (1997) Life and death in the cytoplasm: Messages from the 3' end. *Curr Opin Genet Dev* 7: 220–232.
- Lykke-Andersen J (2001) mRNA quality control: Marking the message for life or death. *Curr Biol* 11: R88–R91.
- Singh G, Lykke-Andersen J (2003) New insights into the formation of active nonsense-mediated decay complexes. *Trends Biochem Sci* 28: 464–466.
- Wagner E, Lykke-Andersen J (2002) mRNA surveillance: The perfect persist. *J Cell Sci* 115: 3033–3038.
- Culbertson MR (1999) RNA surveillance. Unforeseen consequences for gene expression, inherited genetic disorders and cancer. *Trends Genet* 15: 74–80.

64. Nagy E, Maquat LE (1998) A rule for termination-codon position within intron-containing genes: When nonsense affects RNA abundance. *Trends Biochem Sci* 23: 198–199.
65. Glazier AM, Nadeau JH, Aitman TJ (2002) Finding genes that underlie complex traits. *Science* 298: 2345–2349.
66. Korstanje R, Paigen B (2002) From QTL to gene: The harvest begins. *Nat Genet* 31: 235–236.
67. Harries LW, Hattersley AT, Ellard S (2004) Messenger RNA transcripts of the hepatocyte nuclear factor-1alpha gene containing premature termination codons are subject to nonsense-mediated decay. *Diabetes* 53: 500–504.
68. Perrin-Vidoz L, Sinilnikova OM, Stoppa-Lyonnet D, Lenoir GM, Mazoyer S (2002) The nonsense-mediated mRNA decay pathway triggers degradation of most *BRCA1* mRNAs bearing premature termination codons. *Hum Mol Genet* 11: 2805–2814.
69. Murphy CT, McCarroll SA, Bargmann CI, Fraser A, Kamath RS, et al. (2003) Genes that act downstream of DAF-16 to influence the lifespan of *Caenorhabditis elegans*. *Nature* 424: 277–283.